
ПОСТРОЕНИЕ И ВИЗУАЛЬНОЕ ОТОБРАЖЕНИЕ РОДОСЛОВНЫХ СТРУКТУР – ПУТЬ К СОЦИОЛОГИЧЕСКОМУ АНАЛИЗУ¹

И.Н. Боголюбов, О.Б. Божков

(Санкт-Петербург)

Статья посвящена проблеме формирования и поддержки информационной базы родословных структур², используемой для анализа различных социологических макрохарактеристик. Такое назначение базы данных предопределяет ее объемность и унифицированность и, следовательно, предъявляет исключительно высокие требования к простоте и наглядности процедур ввода, контроля и корректировки данных. Непосредственно из специфики использования родословных структур вытекает и требование к наглядности и естественности соответствующих процедур навигации. Разработан и описывается активный графический интерфейс, обеспечивающий как наглядное отображение родословной структуры, так и управление структурой путем синхронизации работы графической сети и таблиц информационной базы.

Ключевые слова: базы данных, биографические события, визуализация, граф, графическое отображение, генеалогия, информационные ресурсы, компьютерная технология, макрохарактеристики социологические, мно-

1 Настоящая работа выполняется при финансовой поддержке РФФИ, грант №98-06 80308 "Методологические и методические принципы компьютерной технологии построения и анализа генеалогических деревьев (структур) в биографических исследованиях".

2 Под родословной структурой мы подразумеваем обобщение понятий генеалогического и родословного деревьев, то есть совокупность всех индивидов, связанных родственно-свойственными отношениями.

жественные (к-местные) отношения, многоуровневая сеть, математическая модель, поколения, родословие, родственные и свойственные отношения, социологические задачи.

Введение

Возросший во всем мире интерес самых разных людей (представителей различных социальных слоев и стран) к собственным историческим корням, к истории своих семей, вызвал бурное развитие компьютерных технологий в этой области. Публикуемые на эту тему материалы и разработки охватывают весьма широкий круг вопросов. Это и стандарты на форматы представления информации [1], и протоколы обмена данными, и собственно информационные ресурсы – базы данных многочисленных конкретных родословных деревьев [2]. Представлена и организационная проблематика – сообщества пользователей и их деятельность, проводимые форумы и т.п. [3]. Но самую значительную группу материалов составляют разработки программного продукта, обеспечивающего построение, сохранение и визуализацию родословных структур [4,5,6]. Оценить объем этой группы затруднительно, поскольку темпы появления все новых разработок или модификации существующих очень высоки. Однако можно сказать, что наиболее значительных и авторитетных программных разработок уже насчитывается более двух десятков.

По аналогии с английским названием разработок этого рода (*Genealogy Software*) будем далее называть их *родословными программами*, имея в виду, что это – не более, чем системы компьютерной поддержки оформления родословных структур.

О популярности, с одной стороны, и о степени проработки этих программ, с другой стороны, свидетельствует тот факт, что уже выработался некий неформальный "стандарт" как на

набор предоставляемых пользователю возможностей, так и на требования к оформлению родословий и к пользовательскому интерфейсу [7].

Среди функциональных возможностей выделим следующие:

- задание широкого спектра разнообразных индивидуальных характеристик и событий (как predetermined, так и произвольных, задаваемых пользователем);
- поиск персонажей по заданному поисковому образу;
- составление списков предков, потомков, браков и некоторых иных фиксированных групп персонажей;
- построение восходящих (собственно родословных) и нисходящих (генеалогических) деревьев;
- поддержка полного набора всех стандартных форматов графических изображений;
- поддержка текстовых форматов RTF и HTML;
- обмен информацией с ресурсами сети Internet;
- ручная настройка элементов оформления семейной истории (рамки, шрифты, цветовые выделения и т.п.);
- сканирование и редактирование фотографий;
- сохранение и выдача аудио- и видеоинформации, демонстрация слайдов со звуковым сопровождением.

Новые требования к построению родословной структуры

Уже из перечня типичных характеристик существующих родословных программ, особенно из его последней части, видно их целевое назначение - индивидуальное оформление с помощью компьютерной технологии истории собственного семейного клана. Очевиден также чисто иллюстративный характер этих программ: они ориентированы на фиксацию характе-

ристик конкретного множества индивидов и их эффективное визуальное отображение.

Стоящая перед нами задача, описанная в [8], выдвигает принципиально иные требования к родословным программам. Мы рассматриваем построение и использование родословной структуры прежде всего как необходимые составляющие специфического метода изучения социологических объектов. При этом хотелось бы подчеркнуть, что мы не случайно употребляем множественное число, говоря о социологических объектах, ибо в их число попадают как единичные родословные структуры, так и их массивы или множества. С этой точки зрения, мы, к сожалению, вынуждены констатировать, что ни одна из многочисленных существующих разработок не имеет сколь-нибудь серьезного инструментария для анализа агрегированных характеристик родословной структуры. Это исключает возможность их непосредственного использования в поставленных исследовательских целях. Более того, даже адаптировать эти разработки для проведения с их помощью социологического анализа родословных структур и их массивов не представляется возможным.

Причина этого в абсолютной закрытости функциональной части разработанных программ (заметим, что в информационной части большинство разработанных систем открыто для внесения специфических изменений в структуру данных). Истины ради необходимо отметить, что даже если бы адаптация этих разработок была возможна, ею, увы, не удалось бы ограничиться – потребовалась бы существенная корректировка ряда блоков, вызванная принципиально иными свойствами решаемых задач.

Таким образом, единственным рациональным путем реализации компьютерной технологии анализа родословных структур оказывается создание оригинального программного обеспечения, учитывающего опыт предшествующих разработок (вос-

произведение удачных решений, корректировка ошибочных и игнорирование решений, неадекватных поставленной задаче).

Характерной особенностью существующих родословных программ является сугубо приватный характер получаемых с их помощью результатов. Построенные родословные структуры представляют всего лишь точку зрения их составителей на то, что и как в ней отобразить. Они не претендуют на объективность, не подвергаются сопоставлению между собой, не рассматриваются как база для научного анализа. При таких условиях для родословных структур допустимы неполнота и противоречивость задания, неунифицированность набора характеристик и событий, описывающих их фигурантов.

При этом, информированность составителя о деталях своей родословной структуры (или возможность получения этих знаний) помогает ему снимать противоречия и заполнять лакуны в описании своего родословия, постепенно формируя все более полную и точную картину.

Совершенно иная ситуация складывается при использовании родословных структур для социологического анализа.

Во-первых, в этом случае информационную базу исследования образует целый массив таких структур, и их описание в целях сопоставимости, возможности типизации и т.п. должно быть определенным образом унифицировано. Следовательно, предоставление составителю возможности произвольного задания характеристик, описывающих персонажей родословной структуры, является не только излишним, но и недопустимым. Набор используемых характеристик полностью определяется задачами предстоящего исследования и задается на этапе настройки программы.

Во-вторых, процесс формирования информационной базы носит сугубо производственный, поточный характер. Исполнитель, заполняющий соответствующую базу данных, информационно и психологически отчужден от того материала, с которым он имеет дело. Особенно, если это не исследователь, кото-

рый, по крайней мере, представляет себе суть предстоящего анализа, а просто технический работник, механически выполняющий рутинную процедуру.

Вытекающее из этой особенности требование – предельно четкая организация ввода исходной информации, исключающая неоднозначность ее интерпретации. В качестве примера игнорирования этого требования можно сослаться на такие солидные программные продукты как *Corel Family Tree* или *Family Tree Maker*, в которых, в частности, ребенок может быть некорректно связан с родителями в случае не единственного брака последних. Если составитель – один из персонажей этой родословной структуры, то, вследствие его достаточной информированности об остальных персонажах его семейного клана, он такой ошибки не допустит. Но при поточном процессе заполнения базы данных это положение недопустимо, так как вероятность ошибки существенно возрастает.

В ходе аналитических исследований в интерактивном режиме пользователь, наблюдая за полученными промежуточными результатами, обычно может оперативно менять стратегию анализа. Важнейшим условием эффективности такого режима является наличие у исследователя целостного представления об исследуемой родословной структуре. Его наиболее наглядной формой является графическое отображение **всей системы взаимосвязей** в структуре. В то же время, ни одна из известных нам родословных программ не дает такого представления, предлагая возможность построения лишь изолированных друг от друга восходящих или нисходящих деревьев. Эти усеченные представления родственных связей, приемлемые в целях иллюстрации, явно недостаточны для решения серьезных исследовательских задач.

Отсюда вытекает третье требование – широкое использование графического представления структуры, обеспечивающее как отображение всей родословной в целом и ее отдельных

фрагментов, так и визуальную управляемую навигацию по всей рассматриваемой структуре.

Реализация модели родословной структуры

Сразу отметим, что мы в нашей модели не будем в точности повторять все приемы и ухищрения, успешно зарекомендовавшие себя в работе с родословными структурами по традиционной технологии (см., например [9]), поскольку последняя принципиально отличается своими возможностями от компьютерной.

В качестве иллюстрации этого замечания сошлемся на два примера. В традиционной технологии, наряду с полной информацией об индивиде, которая сводится в массив *генеалогических досье*, организован частично дублирующий ее массив оперативной информации, составленный из *генеалогических карточек*. Это оправдано, поскольку вызвано значительной экономией труда и времени при обращении к оперативной информации в случаях, когда большего и не требуется. В компьютерной технологии такое организационное эшелонирование информации является излишним.

Еще более наглядный пример связан с проблемой *нумерации персонажей*. В традиционной технологии на нумерацию возложены функции идентификации персонажа по полу, поколению, степени близости к другому персонажу и т.д., что сильно усложняет используемую систему нумерации и все равно решает проблему неудовлетворительно. В компьютерной технологии с ее возможностями быстрой формальной обработки большого объема данных более естественно все эти идентификационные параметры фиксировать по отдельности и анализировать непосредственно, а место каждого персонажа в родословии более наглядно отображать в графическом виде.

С другой стороны, следует адаптировать и применить накопленные действительно удачные приемы и находки, подкре-

пленные не столько традицией и привычкой, сколько целесообразностью их использования и в компьютерной технологии. Это соображение, в первую очередь, относится к форме графического представления родословных структур.

Как указывалось выше, родословная структура представляет собой совокупность всех индивидов, связанных родственно-свойственными отношениями. В качестве математической модели такого рода объектов обычно рассматривается граф, вершины которого представляют индивидов, а ребра - соответствующие отношения между ними. В этом случае вершины должны быть двух типов (в соответствии с полом индивида), а ребра - в соответствии с родственным статусом – как минимум четырех типов (без учета возможных модификаций, о чем речь ниже). Для решения аналитических задач на графе, таких, как определение сечений подграфов, мощности вершин, длин путей между вершинами, анализ связности графа и т.п., эти усложнения не создают никаких трудностей. Однако использование данной модели в качестве формы графического представления объекта не обеспечивает достаточной наглядности (см. примеры на рис. 1а, 1б).

ются понятием "связка", отображающим множественное (k -местное) отношение внутри некоторой группы k вершин. В примере 1в представлено 4-местное отношение (связка) персонажей A, B, C, D , описывающее их как нуклеарную семью, а в примере 1г связка задает сиблингов (см. рис. 1).

Итак, *связка группы вершин – это совокупность всех парных отношений этих вершин.*

Для описания свойств связки введем следующие обозначения:

$I = \{i_1, i_2, \dots, i_k\}$, ($k \geq 2$) – группа k вершин родословной структуры,

$|I|$ – мощность группы I (число ее вершин),

$P(I)$ – связка группы I ,

I_r , ($0 \leq r \leq k-2$) – r -редуцированная группа I (любое подмножество $k-r$ вершин из I),

$R = C_k^r$ – число вариантов r -редуцированной группы,

$I_{r,j}$ – j -й вариант r -редуцированной группы I .

По аналогии с обычными ребрами будем приписывать связке булевы значения: 1, если она существует, и 0 - в противном случае.

Очевидны следующие свойства групп вершин и их связок:

$$I_r \subseteq I \quad (1)$$

$$I_0 \equiv I \quad (2)$$

Вырожденная связка (2-местная связка) содержит в себе единственное ребро и, следовательно, существует тогда и только тогда, когда существует это ребро.

Теорема. *Для существования невырожденной связки $P(I)$ необходимо и достаточно, чтобы существовали все 1-редуцированные связки $P(I_{1j})$.*

Доказательство необходимости.

Предположим обратное – это условие не необходимо. Но тогда отсутствие некоторой связки $P(I_{1j})$ означает отсутствие каких-то ребер в I_r , а значит, согласно свойству (1), и в группе I .

Но тогда не будет существовать $P(I)$. Следовательно, это условие необходимо.

Доказательство достаточности.

Пусть $k = 3$. Тогда все 1-редуцированные связки – это ребра. Из определения связки следует достаточность для этого случая, а по индукции она доказывается и для всех $k > 3$.

Следствие: Определение существования связки может быть выполнено с помощью формулы:

$$P(I) = \& P(I_{1,j}), \quad (1 \leq j \leq R) \quad (3)$$

Всю информацию, определяющую родословную структуру, можно условно разделить на две категории – данные, описывающие собственно индивидов, и данные, описывающие связи между индивидами.

В рамках первой категории можно выделить такие, наличие которых принципиально инвариантно относительно любого конкретного лица, например: фамилия, дата и место рождения, образование, социальный статус (занятие), место жительства и т.п.

Следующая группа данных – события. Это – данные, отражающие динамику развития некоторого процесса. Среди событий целесообразно выделить следующие типы: браки, учеба (повышение квалификации), трудовая деятельность, изменения профессионального статуса, миграции. Заметим, что состав этой группы данных – наименее стабильный из всех данных, описывающих родословную структуру. Он в значительной степени определяется задачами социологического анализа, а также специфическими особенностями исследуемого контингента.

Третья группа данных – описания нестандартизованных характеристик и событий, представляемые, обычно, в развернутой, повествовательной форме (неструктурированная информация). Как правило, они не используются непосредственно в аналитических процедурах, однако, могут применяться для внесения тех или иных уточнений в стандартные характеристики индивидов на этапе формирования информационной базы. Бо-

лее того, они являются основой для корректировки самого набора характеристик, описывающих персонажей родословной структуры.

Наконец, последняя группа данных – описатели родственного статуса, отражающие типы связей между персонажами структуры. Принято различать непосредственные (прямые) и опосредованные (косвенные) связи. Непосредственные связи задаются родственными отношениями типа "родители-дети" и "братья-сестры" (сиблинги), а также свойственными отношениями типа "супруги". Модификации непосредственных связей, принятые в модели, представлены в таблице 1.

Состав этой таблицы, в основном, определяется потенциально возможными родственно-свойственными отношениями и практически остается неизменным. Тем не менее, в зависимости от специфических условий конкретных исследований, число модификаций и их содержание может варьироваться. Так, возможно включение в таблицу крестных и крестников, кормилиц, посаженных родителей и т.д. Таблица построена таким образом, что с ее помощью можно автоматически кодировать "ответный" статус родственно-свойственных отношений.

Пусть персонаж **A** определил персонажа **B** как родственника, имеющего статус *i*. Тогда индекс "ответного" статуса персонажа **A** по отношению к персонажу **B** может быть вычислен по формуле:

$$Q_i + 8 * \text{Mod} + \text{Sex} ,$$

где Sex – пол персонажа **A** (1 – мужской, 0 – женский), а остальные параметры относятся к персонажу **B**.

В соответствии с определением родословной структуры любая пара ее персонажей **A** и **B** связана, причем эта связь может быть представлена посредством цепочки прямых связей между персонажами, лежащими на пути от **A** к **B**. Следовательно, для адекватного задания системы связей в родословной структуре достаточно зафиксировать лишь множество непосредственных связей между ее персонажами.

Это дает возможность обеспечить при вводе данных целостность выстраиваемой сети, то есть отсутствие разрывов, приводящих к возникновению в ней изолированных фрагментов, наличие которых свидетельствует о неполноте представления структуры. Для реализации такого подхода выстраивание структуры осуществляется путем заполнения окрестности любого уже включенного персонажа структуры (будем называть его "активным") индивидами, непосредственно связанными с ним. При этом каждый новый вводимый персонаж расширяет сеть (сведения о нем включаются в описание активного персонажа) и, одновременно, сам встраивается в сеть, включая в свое

описание данные об активном персонаже. Наличие однозначного соответствия "прямого" и "ответного" статуса непосредственных связей, которое определяется структурой таблицы 1, позволяет кодировать родственный статус активного персонажа автоматически.

Если новый персонаж оказывается членом группы, связанной между собой k -местным отношением, то необходимо, во-первых, закодировать статус активного персонажа по отношению ко всем членам этой группы и, во-вторых, вычислить геометрические координаты соответствующих взаимосвязанных графических элементов сети. Для этой цели используется группа специально разработанных суперпозиционных алгоритмов.

Каждый персонаж родословной структуры должен быть однозначно идентифицируемым (как содержательно, так и графически). Этому условию практически всегда удовлетворяет задание следующего набора сведений: *пол; фамилия, имя, отчество; дата рождения; дата смерти*. Тем не менее, на этапе ввода (если предусмотрена последующая процедура дополнения и уточнения) можно ограничиться сокращенным набором идентифицирующих данных (минимальный набор может быть ограничен всего одним идентификатором — характеристикой «пол» персонажа).

Высокая степень наглядности отображения родословной структуры в виде графической сети стимулирует попытку расширить ее функциональное назначение от иллюстрации до управления формированием этой структуры. С этой целью реализован дуплексный режим работы, то есть синхронизирована работа табличной картотеки и графической сети. Все действия, которые осуществляются над картотекой – ввод нового персонажа, исключение персонажа, смена активного персонажа, – немедленно отражаются на структуре сети и, наоборот, изменение конфигурации сети приводит к моментальной реорганизации картотеки.

Никаких ограничений на последовательность ввода персонажей (кроме условия связности сети) не накладывается. Это вызывает необходимость в автоматической корректировке геометрических характеристик сети при переходе к новым, более высоким уровням родословной структуры. С другой стороны, автоматизированное формирование геометрической конфигурации далеко не всегда обеспечивает наибольшую наглядность. Поэтому предусмотрена возможность ручной корректировки сети путем непосредственного перемещения вершин и групп вершин на плоскости чертежа в пределах заданного уровня.

Заключение

Итак, начальный этап пути пройден. Разработана технология формирования и сопровождения информационной базы, описывающей родословную структуру. Ее ключевым принципом является нацеленность на задачи социологического анализа контингентов, представленных этими родословными структурами. Это предполагает необходимость накопления массива данных, значительного по объему и стандартизованного по его информационной структуре. В разработанной компьютерной системе учтены вытекающие из этого обстоятельства требования к полноте и непротиворечивости процедур ввода данных, а также к оперативности и ясности навигации в родословной структуре.

В значительной степени это обеспечивается за счет реализации активного графического интерфейса, дающего не только наглядное отображение родовой структуры, но и инструмент формирования параметров аналитических задач, и его синхронизации с табличной картотекой характеристик индивидов структуры (дуплексный режим).

Ряд результатов выполненной разработки оказывается полезным не только при построении и визуализации родословных структур, но и при их исследовании, в частности, при исследо-

вании различных родственно-свойственных групп (их состав, степень неполноты и т. п.).

Уже сейчас просматривается, какую роль в этом могут сыграть созданная таблица модификаций прямых родственных статусов и обобщение понятия ребра на случай многоместных отношений. Эти соображения дают уверенность в целесообразности дальнейшего развития созданной компьютерной технологии путем наращивания ее аналитического инструментария.

ЛИТЕРАТУРА

1. Best Genealogy Links On the WWW // <http://home.earthlink.net/~middlton/topten.html>
2. Word Family Tree Super Bundles// <http://www.familytreemaker.com/wftsuperbundles/index.html>
3. Internet Directory - Genealogy // <http://www.genealogy.org/directory.shtml>
4. *Nurse H. L.* The ROOTS Story. Software to Make History // <http://www.sonic.net/rstory.html>
5. Latest Version of Family Tree Maker // <http://www.commsoft.com>
6. Corel Family Publisher and PAF Mate // <http://www.genealogy.demon.co.uk/pafmate.htm>
7. Genealogy Software – Features Listed by Developers // <http://www.toltbbs.com/~kbasile/features.htm>
8. Божков О.Б. Родословные (генеалогические) деревья, как источник данных и объект социологического анализа//Социологический журнал, 1998. №3-4.
9. Введение в генеалогию/ Сост. С.В.Белецкий. С.-Пб.: Государственная академия культуры, 1997.

МОДИФИКАЦИИ НЕПОСРЕДСТВЕННЫХ СВЯЗЕЙ

Таблица 1.

	Mod = 0	Mod = 1	Mod = 2	Mod = 3
i	Q_i	Модификация 1	Модификация 2	Модификация 3
0	Базовый статус Мать	Мачеха	Усыновительница	Опекунша
1	Отец	Отчим	Усыновитель	Опекун
2	Жена	Подруга	Сожительница	-
3	Муж	Друг	Сожитель	-
4	Дочь	Падчерица	Удочеренная	-
5	Сын	Пасынок	Усыновленный	-
6	Сестра	Единоутробная	Единокровная	Сводная
7	Брат	Единоутробный	Единокровный	Сводный